

10/7/4  
DIALOG(R) File 347:JAPIO  
(c) 1999 JPO & JAPIO. All rts. reserv.

04952042      \*\*Image available\*\*  
PARALLEL PROCESSING COMPUTERS

PUB. NO.:        07-244642 [JP 7244642 A]  
PUBLISHED:      September 19, 1995 (19950919)  
INVENTOR(s):    OGAMI YASUHIRO  
APPLICANT(s):   SANYO ELECTRIC CO LTD [000188] (A Japanese Company or  
                 Corporation), JP (Japan) .  
                 GIJUTSU KENKYU KUMIAI SHINJIYOUHOU SHIYORI KAIHATSU KIKO  
                 [000000] (A Japanese Company or Corporation), JP (Japan)  
APPL. NO.:      06-034384 [JP 9434384]  
FILED:          March 04, 1994 (19940304)

#### ABSTRACT

PURPOSE: To improve the efficiency of the whole system while reducing mutual interference and maintaining the consistency of file management information.

CONSTITUTION: This parallel processing computers are equipped with secondary storage devices 20-28 that can be accessed by processor groups 12-18, and master secondary storage control units by files are selected and determined among plural secondary storage control units 30-38 by random number and modulo arithmetic. When a file is written, the file data are written in a secondary storage device which is light in load in \*log\* format, and a master secondary storage control unit writes file management information in a secondary storage device in \*log\* format correspondingly. A file management information table in the secondary storage control unit is updated so that the result is reflected. When the file is \*read\* out, the master secondary storage control unit returns the file management information to the processor at the \*request\* source at a \*request\* and the processor accesses desired file \*data\* \*based\* on the file management information.

10/7/5

(19)日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11)特許出願公開番号

特開平7-244642

(43)公開日 平成7年(1995)9月19日

(51)Int.Cl.<sup>6</sup>

G 0 6 F 15/16  
3/06

識別記号

3 5 0 Z  
3 0 1 B

庁内整理番号

F I

技術表示箇所

審査請求 未請求 請求項の数 5 O L (全 9 頁)

(21)出願番号 特願平6-34384

(22)出願日 平成6年(1994)3月4日

(71)出願人 000001889

三洋電機株式会社

大阪府守口市京阪本通2丁目5番5号

(71)出願人 593162453

技術研究組合新情報処理開発機構

東京都千代田区東神田2-5-12 龍角散ビル8階

(72)発明者 大上 靖弘

大阪府守口市京阪本通2丁目5番5号 三洋電機株式会社内

(74)代理人 弁理士 山田 義人

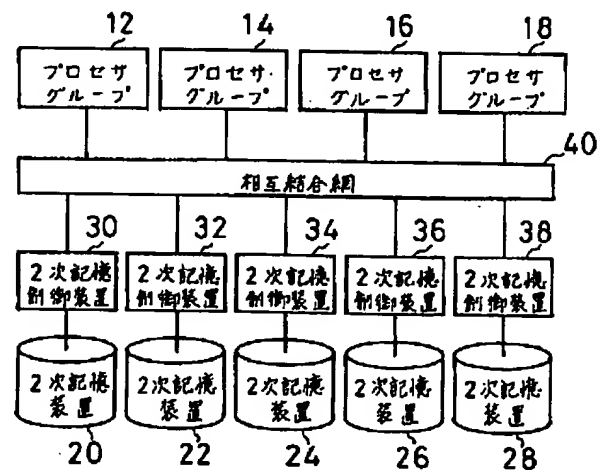
(54)【発明の名称】 並列処理計算機

(57)【要約】

【構成】 各プロセサグループ12~18からアクセス可能な2次記憶装置20~28を備え、乱数およびモジュロ演算によってファイル毎のマスタ2次記憶制御装置を複数の2次記憶制御装置30~38の中から選択して決定する。ファイルの書き込み時には、ファイルデータをログ形式によって負荷の低い2次記憶装置に書き込み、それに応じてマスタ2次記憶制御装置はファイル管理情報をログ形式で2次記憶装置に書き込む。その結果を反映するように2次記憶制御装置内のファイル管理情報テーブルを更新する。ファイルの読み出し時には、マスタ2次記憶制御装置は要求に応じてファイル管理情報を要求元のプロセサに返し、プロセサはそのファイル管理情報に基づいて所望のファイルデータへアクセスする。

【効果】 相互干渉を軽減でき、ファイル管理情報の一貫性を維持でき、システム全体の効率も向上する。

10



## 【特許請求の範囲】

【請求項1】複数のプロセサと、前記複数のプロセサのそれぞれからアクセス可能でありかつそれぞれデータおよびデータ管理情報が格納される複数の2次記憶手段と、それぞれ前記2次記憶手段の少なくとも1つ以上を制御しかつメモリを有する複数の2次記憶制御手段とを備える並列処理計算機において、前記2次記憶手段上に前記データおよび前記データ管理情報をログ形式で格納することによって前記2次記憶手段を動的に選択することを特徴とする、並列処理計算機。

【請求項2】前記データ管理情報が格納される前記2次記憶手段上の位置を示す位置データを管理するマスタ2次記憶制御手段を前記2次記憶制御手段の中から選択する選択手段、および前記2次記憶手段に格納されるデータ管理情報が更新されたとき、対応するマスタ2次記憶制御手段で管理する前記位置データを更新する更新手段を備える、請求項1記載の並列処理計算機。

【請求項3】前記選択手段は、所定の法則に従って前記マスタ2次記憶制御手段となるべき前記複数の2次記憶制御手段のいずれかを選択する、請求項2記載の並列処理計算機。

【請求項4】前記各データ毎のデータ管理情報が存在する前記2次記憶手段上の位置を示す位置データをそれぞれ管理するマスタ2次記憶制御手段を所定の法則に従って前記複数の2次記憶制御手段の中から選択する選択手段、

前記プロセサの要求に応じて対応する前記マスタ2次記憶制御手段から所望のデータ管理情報を前記プロセサに与える手段、および得た前記データ管理情報に基づき前記プロセサが所望のデータへアクセスする手段を備える、請求項1記載の並列処理計算機。

【請求項5】前記2次記憶手段に格納されるデータ管理情報が更新されたとき、対応するマスタ2次記憶制御手段で管理する前記位置データを更新する更新手段を備える、請求項4記載の並列処理計算機。

## 【発明の詳細な説明】

## 【0001】

【産業上の利用分野】この発明は並列処理計算機に関し、特にたとえば複数のプロセサと複数の2次記憶制御装置によって制御される複数の2次記憶装置とを有する並列ファイルシステムなどに用いられる、並列処理計算機に関する。

## 【0002】

【従来の技術】複数のプロセサを有する並列処理計算機において、計算に必要な大量のデータに対する高速なアクセスを可能にするために、2次記憶装置を複数個並列に接続したシステムが提案されている。このような並列処理計算機はマルチユーザ、マルチプロセスでの使用が不可欠であり、異なるユーザの複数のプロセスが混在す

る使用環境が想定される。

## 【0003】

【発明が解決しようとする課題】しかし、このような使用環境においては、複数のプロセスが相互結合網や主記憶装置や2次記憶装置などの共有の資源を利用するために、資源獲得の競合によって性能低下が起こり得る。このような性能低下を防ぐためには、共有資源獲得の競合によるプロセス間の相互干渉を排除することが重要となる。

【0004】相互結合網や主記憶装置に関しては、パーティションによる領域分割などによってプロセス間の相互干渉の排除が可能である。しかし、2次記憶装置上のデータは本質的に全プロセスの共有資源であるため、相互干渉の排除が困難である。それゆえに、この発明の主たる目的は、相互干渉を軽減できる、並列処理計算機を提供することである。

【0005】この発明の他の目的は、ファイル管理情報の一貫性の維持が容易になる、並列処理計算機を提供することである。この発明のその他の目的は、システム全体の効率が向上する、並列処理計算機を提供することである。

## 【0006】

【課題を解決するための手段】この発明は、複数のプロセサと、複数のプロセサのそれぞれからアクセス可能でありかつそれぞれデータおよびデータ管理情報が格納される複数の2次記憶手段と、それぞれ2次記憶手段の少なくとも1つ以上を制御しかつメモリを有する複数の2次記憶制御手段とを備える並列処理計算機において、2次記憶手段上にデータおよびデータ管理情報をログ形式で格納することによって2次記憶手段を動的に選択することを特徴とする、並列処理計算機である。

## 【0007】

【作用】2次記憶手段へのデータおよびデータ管理情報の格納をログ形式で行うので、データおよびデータ管理情報の更新時にはそれらの書き込み場所は自由である。したがって、プロセス間の相互干渉を避けるようにデータの書き込み先である2次記憶手段が動的に選択される。

【0008】また、選択手段によってデータ毎にマスタ2次記憶制御手段を選択する。そして、或るデータおよびそのデータ管理情報はログ形式で書き込まれるので、更新する毎にデータ管理情報が書き込まれる2次記憶手段上の位置が変わり、場合によってはデータ管理情報が書き込まれる2次記憶手段も変わる可能性がある。データ毎のデータ管理情報の格納位置が変わることに伴って、対応するマスタ2次記憶制御手段で管理する位置データを更新手段で更新する。

【0009】このとき、さらに、データ毎のマスタ2次記憶制御手段をたとえば乱数を用いて偏らないように全ての2次記憶制御手段の中から選択してもよい。また、

10

20

30

40

50

マスタ2次記憶制御手段を、たとえば乱数を発生したりして、所定の法則に従って偏らないように選択した場合、プロセッサがデータを要求すると、データ毎に対応するマスタ2次記憶制御手段から、データ管理情報をそのプロセッサに与える。プロセッサはこのデータ管理情報に基づいて所望の2次記憶手段に対してアクセスし、所望のデータを得る。

#### 【0010】

【発明の効果】この発明によれば、プロセス間の相互干渉を避けるようにデータおよびそのデータ管理情報をログ形式で書き込むことによって、特に書き込み動作における相互干渉を軽減できる。また、データ管理情報の位置データを、データ毎のマスタ2次記憶制御手段内で管理するので、ファイル管理情報の格納位置が変更しても対応するマスタ2次記憶制御装置内の位置データを書き換えるだけでよいので、容易にファイル管理情報の一貫性を維持できる。

【0011】さらに、マスタ2次記憶制御手段を特定の2次記憶制御手段に偏らないように選択すると、データ管理情報の位置データの格納・更新やデータ管理情報に対するアクセスのための負荷が偏ることを避けることができ、システム全体の効率が向上する。この発明の上述の目的、その他の目的、特徴および利点は、図面を参照して行う以下の実施例の詳細な説明から一層明らかとなる。

#### 【0012】

【実施例】図1に示すこの実施例の並列処理計算機は、たとえば並列ファイルシステムとして用いられる。以下、並列処理計算機10を並列ファイルシステムとして用いた場合について説明する。並列処理計算機10はプロセッサグループ12、14、16および18を含む。各プロセッサグループ12ないし18は、それぞれ複数のプロセッサを含む。また、各プロセッサはCPUとメモリとを含む。各プロセッサグループ12ないし18上でそれぞれ1つのプロセスが動作する。プロセスが2次記憶装置20、22、24、26または28にアクセスする場合、実際にはそのプロセスが動作しているプロセッサグループ内の1つのプロセッサによってアクセスが実行される。2次記憶制御装置30、32、34、36および38は、それぞれ2次記憶装置20、22、24、26および28を制御する。注目すべきは、後述するように各2次記憶装置30ないし38はそれぞれメモリを有し、特定のファイルのファイル管理情報テーブルをそのメモリ内で管理することである。ここで、2次記憶装置20ないし28はそれぞれたとえばディスクなどからなり、2次記憶制御装置30ないし38はそれぞれたとえばディスクコントローラなどからなる。また、各プロセッサグループ12、14、16および18と2次記憶制御装置30、32、34、36および38とは、相互結合網40によって相互に結合される。

【0013】このような並列処理計算機10において、ファイル識別番号の生成は図2に示すように行われる。このとき、プロセッサグループ12ないし18に含まれるそれぞれのプロセッサで行われる図2の動作（プロセス）によってファイル識別番号を生成する。図2のステップS1においてファイル識別番号の生成要求があると、最初にファイル毎にマスタ2次記憶制御装置の選択を行う。そのために、ステップS3において乱数を発生し、ステップS5において、 $(\text{マスタ2次記憶制御装置番号} = \text{乱数} \bmod \text{2次記憶制御装置数})$ を計算して、マスタ2次記憶制御装置番号を決定する。このように、マスタ2次記憶制御装置の選択は乱数を使って行うため、統計的には各2次記憶制御装置30ないし38が均等に選択される。したがって、ファイル毎のマスタ2次記憶制御装置が特定の2次記憶制御装置に偏ることを防ぐことができる。このように、ステップS3およびS5が、乱数+モジュロ演算のような所定の法則に従って偏りなくマスタ2次記憶制御装置として特定の2次記憶制御装置を選択する。ここで、各マスタ2次記憶制御装置は、対応するファイル毎の位置データを格納する。位置データは、データ管理情報が格納される2次記憶制御装置上の位置（たとえばアドレス）を示す。

【0014】次いで、ステップS7において、選択された2次記憶制御装置すなわちマスタ2次記憶制御装置との通信を行い、マスタ2次記憶制御装置内のローカルなファイル番号（図6に示す）を獲得する。そしてステップS9において最後に、マスタ2次記憶制御装置の番号と2次記憶制御装置内のローカルなファイル番号とを組み合わせ、図3に示すようなファイル識別番号を生成する。図3に示すファイル識別番号は32ビットで構成され、29ビットの2次記憶制御装置内のローカルなファイル番号と3ビットのマスタ2次記憶制御装置番号とを含む。そして、このファイル識別番号は各2次記憶装置20ないし28内に格納される。このファイル識別番号によってファイルおよびマスタ2次記憶制御装置を一義的に決定できる。

【0015】次いで、ファイル識別番号を用いたファイルアクセスを説明する。まず、図4および図5を参照して、ファイルの書き込みを説明する。ここで、ファイルは、ファイルデータとファイルデータの2次記憶装置上の位置（位置データ）を管理するファイル管理情報とによって構成され、2次記憶装置内に格納される。

【0016】図4に示すステップS11において、プロセスは負荷の低い2次記憶装置を制御している2次記憶制御装置に対してファイルデータの書き込み要求を行い、ファイルデータを書き込む。図5に示すように、たとえば2次記憶装置24の負荷が低ければ、2次記憶制御装置34に対してファイルデータの書き込み要求を行う（図5①）。負荷の低い2次記憶装置の検出は、2次記憶装置20ないし28にそれぞれ接続されている2次

記憶制御装置30ないし38が周期的に2次記憶装置20ないし28の負荷状態を検出し、それを各プロセサグループ12ないし18に知らせることによって、各プロセサ12ないし18は負荷の低い2次記憶装置を知ることができる。

【0017】そして、ファイルデータの書き込みが終了すると、図4に示すステップS13において2次記憶制御装置は書き込まれたファイルデータの2次記憶装置上の位置をプロセスに返す。図5でいえば、2次記憶装置24へのファイルデータの書き込みが終了すると、2次記憶制御装置34は書き込まれたファイルデータの2次記憶装置24上の位置をプロセサグループ12内のプロセサに返す。次いで図4に示すステップS15において、プロセスは、そのファイルのマスタ2次記憶制御装置に対して、ファイル管理情報の書き込み要求を行う。図5でいえば、マスタ2次記憶制御装置となる2次記憶制御装置30に対して書き込み要求を行う(図5③)。そして、図4に示すステップS17において、マスタ2次記憶制御装置は2次記憶装置に、ログ形式でファイル管理情報を書き込む。図5でいえば、マスタ2次記憶制御装置30が2次記憶装置20に、ログ形式でファイル管理情報を書き込む(図5④)。そして、図4に示すステップS19において、その結果を反映するように2次記憶制御装置内のファイル管理情報テーブル(図6)を更新する。図5でいえば、2次記憶装置20内のファイル管理情報の位置の変更を反映するように2次記憶制御装置30内のファイル管理情報テーブルを更新する(図5⑤)。

【0018】ここで、図6に示すファイル管理情報テーブルは、マスタ2次記憶制御装置内のローカルなファイル番号と、そのファイル管理情報の2次記憶装置上の位置との対応表である。すなわち、ファイル管理情報テーブルには、その2次記憶制御装置が管理するファイルのファイル管理情報が、いずれの2次記憶上のいずれの位置に格納されているかを示す位置データを格納する。図6に示す場合、ファイル番号1のファイル管理情報は、2次記憶装置20のアドレス「100」に格納されており、ファイル番号2のファイル管理情報は、2次記憶装置22のアドレス「205」に格納されていることなどを示す。なお、ファイル管理情報テーブルは各2次記憶制御装置内にそれぞれ存在する。

【0019】次いで、図7および図8を参照して、ファイルの読み出しを説明する。まず、図7に示すステップS21において、プロセスはマスタ2次記憶制御装置に対してファイル管理情報の読み出しを要求する。図8でいえば、プロセサグループ12内のプロセサが2次記憶制御装置30に対して読み出し要求を行う(図8①)。図7に示すステップS23において、ファイル管理情報の読み出し要求を受けた2次記憶制御装置はファイル管理情報テーブルからファイル管理情報の2次記憶装置上

の位置を調べ、ファイル管理情報を読み出して、要求元のプロセスに返す。図8でいえば、2次記憶制御装置30が、そのファイル管理情報テーブルからファイル管理情報の2次記憶装置20ないし28上の位置を調べ、ファイル管理情報を読み出してプロセサグループ12内のプロセサに返す(図8②)。そして、図7に示すステップS25において、プロセスは得られたファイル管理情報に基づき、ファイルデータの読み出し(アクセス)要求を各2次記憶制御装置に対して行う。図8でいえば、プロセサグループ12内のプロセサはファイル管理情報に基づき、各2次記憶装置20ないし28に対してファイルデータのアクセス要求を行う(図8③)。

【0020】ここで、図7のステップS21で行われるファイル管理情報の読み出し要求は、ファイル識別番号を用いて行われる。ファイル識別番号の管理等に関しては、UNIXのファイルシステムとほぼ同様である。すなわち、ファイルはファイル名ではなくファイル識別番号で管理され、ファイル名とファイル識別番号との対応関係は任意の2次記憶装置内に書き込まれている。したがって、プロセスは2次記憶装置の中から所望のファイル名に対応するファイル識別番号を検索し、そのファイル識別番号によってマスタ2次記憶制御装置を特定する。したがって、プロセスはファイル識別番号に基づいてマスタ2次記憶制御装置を特定し、ファイル管理情報の内容を得ることができる。

【0021】このような並列処理計算機10において、まず注目すべきはファイルデータおよびファイル管理情報をログ形式で2次記憶装置20ないし28のいずれかに格納していくことである。図9を参照して、ログ形式でのファイルデータの格納について説明する。通常のファイルシステムにおいては、ファイルデータの位置はファイル生成時に決められ、図9(A)に示すように、ファイルデータの更新はその位置で行われる。この場合、1と6とが修正されている。

【0022】これに対して、ログ形式のファイルシステムでは、図9(B)に示すように、ファイルデータの更新をファイルデータの末尾への追加書き込みの形で行う。すなわち、ログとは、2次記憶装置上に存在するデータの変更の履歴である。ログ形式でファイルデータを書き込むことによって、書き込みの高速化が図れる。この一因として、書き込みの際にディスクヘッドの SEEK が不要となることが挙げられる。すなわち、通常のファイルシステムの場合は、データ1'を書き込んだ後、データ6'の位置までディスクヘッドを移動してからデータ6'を書き込むが、ログ形式のファイルシステムの場合は、このディスクヘッドの移動がいらない。

【0023】しかし、図9(B)でもわかるように、データの更新後も古い無効なデータ1と6とが2次記憶装置上に存在するため、これを削除して、その領域を再利用するためのクリーニングという作業が必要となる。こ

れは、プログラミング言語におけるガーベージコレクションと類似した作業である。次いで、「更新されるデータは、元の位置と異なる場所に書き込まれる」というログ形式の特徴を利用して、2次記憶装置を動的に選択するという点について説明する。

【0024】図10に示すように、2次記憶装置AとBとがあるとする。並列処理計算機ではたくさんのプロセッサが並列に動作するので、或るプロセッサが2次記憶装置A上のデータ1と6とをデータ1'と6'とに更新しようとしたとき、他のプロセッサが2次記憶装置A上のデータ3を読み出し中であったという状況が考えられる。この場合、通常の並列ファイルシステムでは、図10(A)に示すように2次記憶装置Aが空くまでデータ1'と6'との書き込みを待たなければならない。

【0025】しかし、ログ形式の並列ファイルシステムでは、図10(B)に示すように、データ1'と6'とを2次記憶装置Aに書き込む必要はないので、2次記憶装置Aが使用中であれば、空いている2次記憶装置Bにデータ1'と6'とを書き込むことができる。このように、2次記憶装置Aが空くまでデータ1'と6'との書き込みを待つ必要がなくなるので、その分書き込みが高速になる。

【0026】このようにこの実施例では、負荷の低い2次記憶装置を動的に選択することによって、2次記憶装置への書き込みに関する、他のプロセッサとの相互干渉を排除することができる。したがって、書き込みの高速化が実現する。続いて、この並列処理計算機10の他の注目すべき点は、ファイル管理情報の一貫性の維持が容易となることである。この点について以下説明する。

【0027】まず、並列ファイルシステムへのファイルデータおよびファイル管理情報の書き込みをログ形式にするだけでは、以下のような不都合が考えられる。すなわち、ログ形式の並列ファイルシステムでは、ファイル管理情報もファイルデータと同様にログ形式で2次記憶装置に書き込まれる。そして、最新のファイル管理情報をプロセッサ側がそのメモリ上で管理する。そして、ログ形式の並列ファイルシステムでは、更新されるファイルデータを負荷の低い2次記憶装置に書き込んでいくので、図11および図12に示すように複数のファイル管理情報が存在することになる。したがって、最新のファイル管理情報を全プロセッサが知っている必要がある。すなわち、ファイル管理情報の一貫性の維持が本質的に必要となる。なお、図11および図12において、inodeはファイル管理情報を示す。

【0028】ここで、たとえば、最新のファイル管理情報をプロセッサ側がメモリ上で管理すると、全プロセッサが同じ情報をもたなければならない。共有メモリやコピーレントキャッシュがあれば問題はないが、並列処理計算機のように数千、数万のプロセッサを有するシステムにおいては共有メモリなどの実装が困難である。共有メモリ

などのないシステムの場合は、全プロセッサがローカルなメモリ上にファイル管理情報のコピーを持つことになるので、ファイル管理情報が更新されるたびに、プロセッサ間で通信を行って、常に最新のファイル管理情報を全プロセッサが持つようにしなければならない。したがって、ファイルの書き込みの効率が低下する。

【0029】そこで、この実施例の並列ファイルシステムでは、2次記憶制御装置がファイル管理情報を管理する。しかし、ファイル管理情報を2次記憶制御装置上のメモリで管理すると、メモリのサイズには限界があるので、管理できるファイルの数は制限されてしまう。そこで実際には、ファイル管理情報をもファイルデータと同様にログ形式で2次記憶制御装置に書き込み、2次記憶制御装置は「ファイル管理情報の2次記憶装置上での格納位置を示す位置データ」をそのメモリ上で管理することにする。

【0030】以下、図13を参照して、より具体的に説明する。ファイルデータおよびファイル管理情報の更新前の状態を図13(A)に示し、更新後の状態を図13(B)に示す。図13(A)に示す更新前では、ファイルX1のファイルデータaないしfは、2次記憶装置Aのアドレス(ブロック番号)101ないし106に格納されている。そして、2次記憶装置Aのアドレス107(斜線部分)には、ファイル管理情報が格納される。ファイル管理情報としては、テーブル42aに示すようにA-101ないしA-106の管理情報が格納される。ここで、ファイルX1の「X」はいずれの2次記憶制御装置であるかを示し、「1」はマスタ2次記憶制御装置内のローカルなファイル番号を示す。A-101の「A」はいずれの2次記憶装置であるかを示し、「101」はアドレスを示す。以下、図13および図14において同様である。

【0031】このとき、ファイル管理情報は2次記憶装置Aのアドレス107に格納されているので、ファイル管理情報の格納位置である位置データを管理する2次記憶制御装置Xのメモリには、ファイル管理情報テーブル44が格納される。ファイル管理情報テーブル44には、マスタ2次記憶制御装置X内のローカルなファイル番号が「1」(ファイルX1)のファイル管理情報が、2次記憶装置A上のアドレス107の位置に格納されているという位置データが格納される。

【0032】そして、図13(B)に示すように、ファイルX1のファイルデータaとeとが、それぞれa'とe'とに変更されたとする。この場合、ファイルデータa'およびe'は、それぞれ負荷の低い2次記憶装置B内のアドレス153および154に格納されている。それに応じて更新されたファイル管理情報が、2次記憶制御装置Yに制御される2次記憶装置Bのアドレス155に格納される。このファイル管理情報には、テーブル42bに示すように、ファイルデータa'はB-153に

格納され、ファイルデータe'はB-154に格納されていることが示される。この点がテーブル42aと異なる。そして、それに応じて2次記憶制御装置Aのファイル管理情報テーブル44は、ファイルX1のファイル管理情報が2次記憶装置B上のアドレス155に格納されていることを示すように、その位置データを書き換える。

【0033】このように、「ファイル管理情報の位置データ」を2次記憶制御装置Xで管理することによって、データが更新されかつそれに従ってファイル管理情報の格納位置が変更しても、2次記憶制御装置X内のファイル管理情報テーブルの一部(位置データ)を書き換えるだけで足りるので、ファイル管理情報の一貫性の維持が容易となる。したがって、従来では、複数のプロセッサ間でのデータ共有のための機構たとえば共有メモリやコヒーレントキャッシュなどのような機構をサポートしないシステムでは、ファイル管理情報の一貫性の維持が困難であったが、このようなシステムであってもこの実施例を適用することによって、少ないオーバーヘッドによって一貫性の維持が容易となる。

【0034】また、たとえば全てのファイルのファイル管理情報の位置データを1つの2次記憶制御装置によって管理すると、一貫性の問題は解決されるが、ファイル管理情報の位置データの入力や更新またはファイル管理情報に対するアクセスがこの2次記憶制御装置だけに偏ることになる。すなわち、1つの2次記憶制御装置に負荷が偏り、システム全体の効率が低下する。したがって、ファイル管理情報の位置データを特定の2次記憶制御装置に偏らないように全ての2次記憶制御装置に分散して管理する。

【0035】具体的には、ファイル識別番号生成時にたとえば乱数を用いて各ファイル毎にマスタ2次記憶制御装置を決定し、各ファイルのファイル管理情報の位置データを対応するマスタ2次記憶制御装置によって管理する。言い換えれば、2次記憶制御装置は、自分がマスタ2次記憶制御装置となっているファイルのファイル管理情報の位置データだけを管理する。

【0036】このように、ファイル管理情報の位置データを全ての2次記憶制御装置に分散して管理するので、1つの2次記憶制御装置に負荷が偏ることはなく、システム全体の効率が向上する。この場合、或る時点での2次記憶装置A、BおよびCならびに2次記憶制御装置X、YおよびZの状態は、たとえば図14のようになる。ここでは、ファイルX1、Y1、Z1、X2が存在し、ファイルX1とX2とのマスタ2次記憶制御装置は2次記憶制御装置Xであり、ファイルY1のマスタ2次記憶制御装置は2次記憶制御装置Yであり、ファイルZ1のマスタ2次記憶制御装置は2次記憶制御装置X2であるとする。なお、2次記憶装置の図中の( )はファイル番号を示す。

【0037】図14において、ファイルX1のファイル管理情報またはファイルX2のファイル管理情報を得ようとするれば、プロセッサは2次記憶制御装置Xにアクセスすればよく、ファイルY1のファイル管理情報を得ようとするれば、プロセッサは2次記憶制御装置Yにアクセスすればよく、ファイルZ1のファイル管理情報を得ようとするれば、プロセッサは2次記憶制御装置Zにアクセスすればよい。このように2次記憶制御装置へのアクセスが分散するので、システム全体の効率が向上する。

【0038】また、各マスタ2次記憶制御装置毎に個々のファイルの属性としてファイル識別番号を付すので、このファイル識別番号に基づいて、ファイルアクセス時に必要なファイル管理情報の位置データを管理するマスタ2次記憶制御装置を容易に知ることができる。なお、2次記憶制御装置は、複数の2次記憶装置を制御するものであってもよい。

【図面の簡単な説明】

【図1】この発明の一実施例を示すブロック図である。

【図2】この実施例のファイル識別番号の生成動作を示すフロー図である。

【図3】ファイル識別番号の一例を示す図解図である。

【図4】ファイルの書き込み動作を示すフロー図である。

【図5】ファイルの書き込み動作の一例を説明するための図解図である。

【図6】ファイル管理情報テーブルの一例を示す図解図である。

【図7】ファイルの読み出し動作を示すフロー図である。

【図8】ファイルの読み出し動作の一例を説明するための図解図である。

【図9】ログ形式でのファイルデータの格納を説明するための図解図である。

【図10】2次記憶装置の動的な選択動作を説明するための図解図である。

【図11】ファイル管理情報の格納状態を示す図解図である。

【図12】各2次記憶装置へファイル管理情報を格納する状態を示す図解図である。

【図13】ファイル管理情報テーブルの更新を説明するための図解図である。

【図14】或る時点での2次記憶制御装置および2次記憶装置の状態を示す図解図である。

【符号の説明】

10 …並列処理計算機

12, 14, 16, 18 …プロセッサグループ

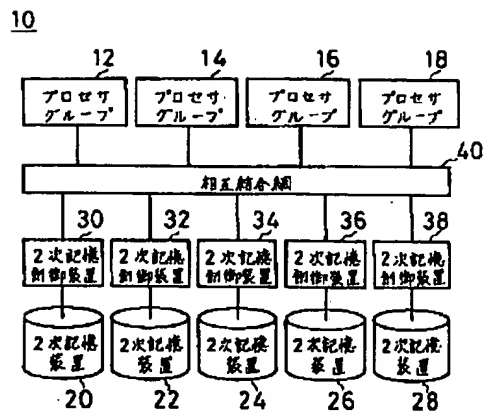
20, 22, 24, 26, 28 …2次記憶装置

30, 32, 34, 36, 38 …2次記憶制御装置

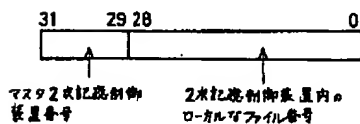
40 …相互結合網

50 44 …ファイル管理情報テーブル

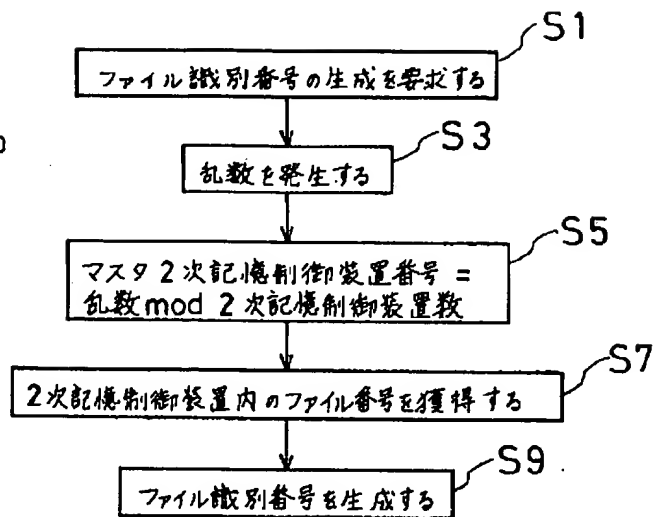
【図1】



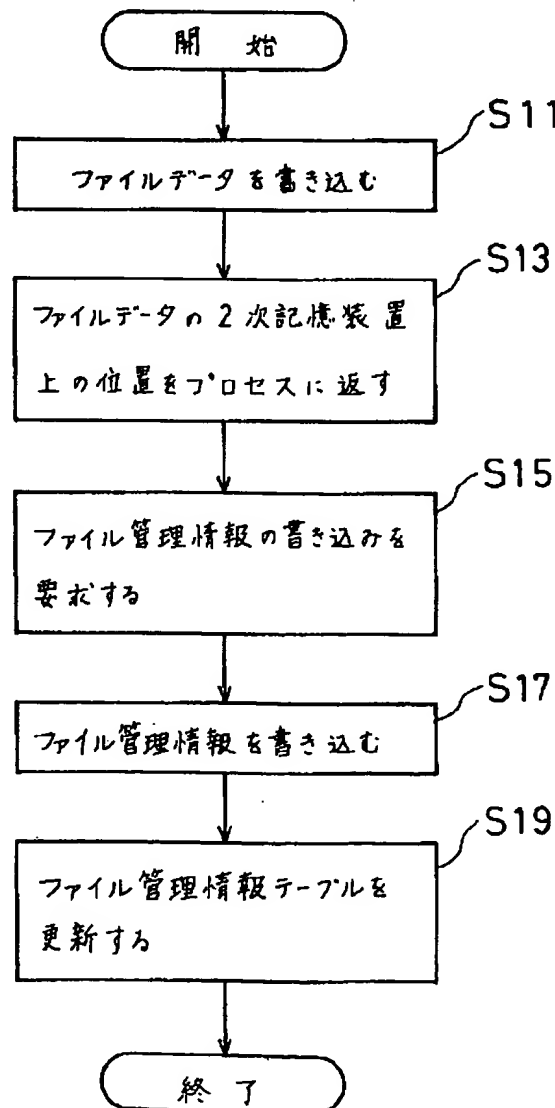
【図3】



【図2】



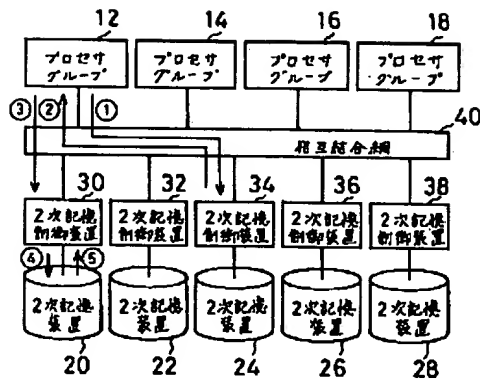
【図4】



【図12】



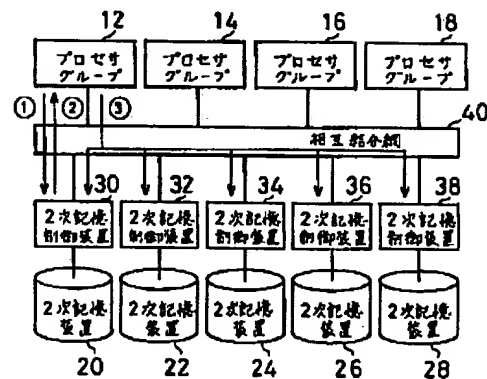
【図5】



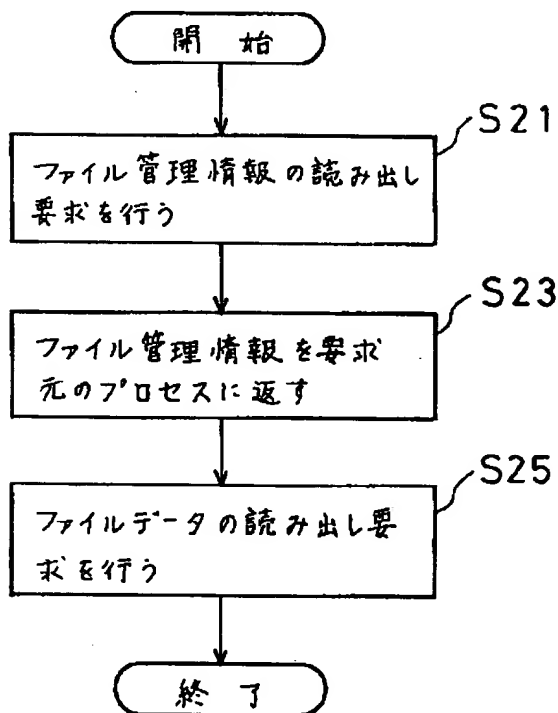
【図6】

ファイル管理情報テーブル	
マスタ2次記憶制御装置内のローカルファイル番号	ファイル管理情報の2次記憶装置上の位置
1	20 - 100
2	22 - 205
⋮	⋮

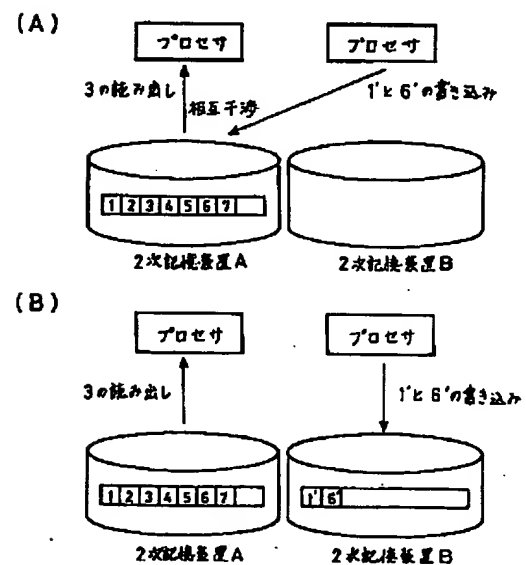
【図8】



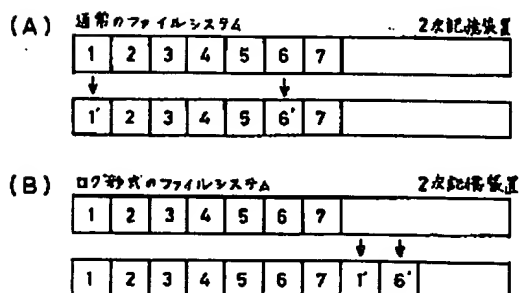
【図7】



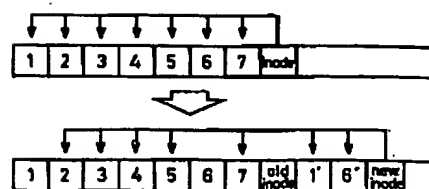
【図10】



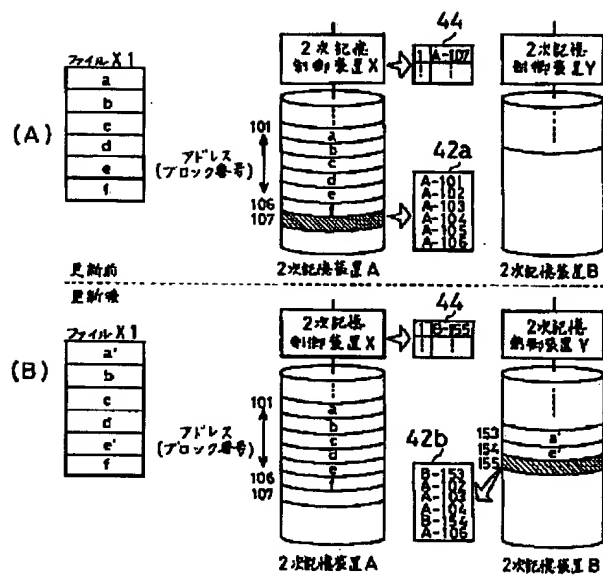
【図9】



【図11】



【図13】



【図14】

